



# Amino Acid Metabolic Origin as an Evolutionary Influence on Protein Sequence in Yeast

## Citation

de Bivort, Benjamin L., Perlstein, Ethan O. Perlstein, Sam Kunes, and Stuart L. Schreiber. 2009. Amino acid metabolic origin as an evolutionary influence on protein sequence in yeast. *Journal of Molecular Evolution* 68(5): 490-497.

## Published Version

doi:10.1007/s00239-009-9218-5

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4454163>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Amino Acid Metabolic Origin as an Evolutionary Influence on Protein Sequence in Yeast

Benjamin L. de Bivort · Ethan O. Perlstein ·  
Sam Kunes · Stuart L. Schreiber

Received: 2 September 2008 / Accepted: 19 February 2009 / Published online: 9 April 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** The metabolic cycle of *Saccharomyces cerevisiae* consists of alternating oxidative (respiration) and reductive (glycolysis) energy-yielding reactions. The intracellular concentrations of amino acid precursors generated by these reactions oscillate accordingly, attaining maximal concentration during the middle of their respective yeast metabolic cycle phases. Typically, the amino acids themselves are most abundant at the end of their precursor's phase. We show that this metabolic cycling has likely biased the amino acid composition of proteins across

the *S. cerevisiae* genome. In particular, we observed that the metabolic source of amino acids is the single most important source of variation in the amino acid compositions of functionally related proteins and that this signal appears only in (facultative) organisms using both oxidative and reductive metabolism. Periodically expressed proteins are enriched for amino acids generated in the preceding phase of the metabolic cycle. Proteins expressed during the oxidative phase contain more glycolysis-derived amino acids, whereas proteins expressed during the reductive phase contain more respiration-derived amino acids. Rare amino acids (e.g., tryptophan) are greatly overrepresented or underrepresented, relative to the proteomic average, in periodically expressed proteins, whereas common amino acids vary by a few percent. Genome-wide, we infer that 20,000 to 60,000 residues have been modified by this previously unappreciated pressure. This trend is strongest in ancient proteins, suggesting that oscillating endogenous amino acid availability exerted genome-wide selective pressure on protein sequences across evolutionary time.

Benjamin L. de Bivort and Ethan O. Perlstein have contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00239-009-9218-5) contains supplementary material, which is available to authorized users.

B. L. de Bivort (✉)  
Rowland Institute at Harvard, Harvard University,  
100 Edwin Land Boulevard, Cambridge, MA 02142, USA  
e-mail: bivort@fas.harvard.edu

B. L. de Bivort · S. Kunes  
Department of Molecular and Cellular Biology, Center for Brain  
Science, Harvard University, 7 Divinity Avenue, Cambridge,  
MA 02138, USA

E. O. Perlstein  
Carl Icahn Laboratory, Lewis-Sigler Institute for Integrative  
Genomics, Princeton University, Princeton, NJ 08544, USA

S. L. Schreiber  
Howard Hughes Medical Institute, Broad Institute of Harvard  
and MIT, 7 Cambridge Center, Cambridge, MA 02142, USA

S. L. Schreiber  
Department of Chemistry and Chemical Biology, Harvard  
University, 12 Oxford Street, Cambridge, MA 02138, USA

**Keywords** Amino acid sequence · Metabolism ·  
Optimization · Oxidation · Protein function · Reduction ·  
Yeast metabolic cycle

## Introduction

Many selective pressures alter the primary amino acid sequence of proteins: metabolic availability of biosynthetic precursors (Tu et al. 2005), hydrophobicity and other physicochemical properties of amino acids (Lobry and Gautier 1994), and codon use bias (Porter 1995), among others. A number of groups have previously studied amino

acid compositional biases using genomic data from many organisms (Pascal et al. 2005; Tekai and Yeramian 2006; Thioulouse and Lobry 1995; Nilsson et al. 2005). Amino acid residues that engender specific catalytic activity or constitute protein–protein interaction domains necessary for function are often conserved during evolution, and when these residues mutate to other amino acids they do so in a chemically conservative manner that reflects selection.

In contrast, amino acid residues that may be substituted during evolution without detriment to protein function per se do not necessarily drift at random. These two examples broadly define the conceptual framework for considering functional constraints on protein sequence. One category of constraints arises passively because of the structure of the genetic code, reflecting genomic G + C content and codon use biases. A number of other constraints on the primary amino acid sequence of proteins fall under the broad heading of biosynthetic cost minimization. They are actively governed by selection and influence amino acid mutability independently of specific protein function. These include translation efficiency, amino acid metabolic cost (Akashi and Gojobori 2002; Seligmann 2003), and depletion of cognate amino acids from the enzymes of amino acid-biosynthetic pathways. Several groups—including one study that documented the selective depletion of sulfur-containing amino acids in highly expressed proteins of a microorganism that dwells in a sulfur-limited environment (Mazel and Marliere 1989), a follow-up study that performed an analysis of intragenomic cognate depletion in carbon-, nitrogen- and sulfur-assimilatory pathways (Baudouin-Cornu et al. 2001), and two independent studies that focused specifically on amino acid biosynthetic pathways in multiple genomes (Alves and Savageau 2005)—have examined the latter phenomenon.

In the last example, the amino acid produced by a pathway of enzymes was selectively eliminated from the primary amino-acid sequences of the self-same enzymes. That observation was made possible by comparing the frequency of each amino acid in a group of functionally related proteins to its proteomic frequency. However, it would be difficult, if not impossible, to determine systematically the extent to which a given amino acid in a protein is affected by “local” pressures on protein function versus “global” pressures imposed by cellular bioenergetics. In an attempt to understand the effects of cellular bioenergetics on amino acid composition, we generalized a previous effort (Perlstein et al. 2007) that detected specific amino acid composition bias in proteins involved in the same cellular function, reasoning that we would detect trends in amino acid composition operating at a systems level, compared with the idiosyncratic functional requirements of individual proteins, by grouping proteins that participate in a common cellular function.

We first show that groups of functionally related proteins have highly variable biases with respect to amino acid composition and that this variability reflects three different amino acid properties: polarity, metabolic cost, and the metabolic origin of the amino acid’s precursors. We then show that this last property (metabolic origin) is uniquely correlated with the different amino acid compositional biases observed in periodically expressed proteins. We finally propose that the intrinsic fluctuations in amino acid concentrations of the yeast metabolic cycle (YMC) (Tu et al. 2005; Richard 2003), which entails alternating oxidative and reductive metabolic phases, has biased protein sequence with respect to all amino acids during evolutionary timescales.

## Materials and Methods

GO gene annotations were downloaded for each species from the European Bioinformatics Institute (<http://www.ebi.ac.uk/GOA/proteomes.html>). Protein sequences were downloaded from the Swiss Institute of Bioinformatics (<http://expasy.org/sprot/>). These data files were imported into MATLAB (MathWorks, Natick, MA) using custom scripts. Amino acid composition biases were calculated across GO-defined groups of proteins by pooling the amino acids from all proteins in the group and then determining the overrepresentation or underrepresentation of each amino acid compared with the proteome-wide expectation. These biases were clustered across GO-defined groups using the UPGMA (Sneath and Sokal 1973) distance method, which is based on a metric in which the distance between GOs is the Euclidean distance between the bias profiles of each GO. Clusters across amino acids were calculated using the Fitch–Margoliash (FM) method (Fitch and Margoliash 1967) based on a Euclidean distance metric. Because GO and amino acid profiles are not related by respective evolutionary processes, and the clustering was largely for visualization, distance methods are appropriate. FM was our preferred clustering algorithm because it assumes nonultrametric data and uses branch swapping to locally optimize trees. UPGMA was used to cluster GO biases because FM was computationally prohibitive for more than approximately 100 terminals. All analyses and statistical tests were performed using MATLAB using built-in or custom scripts.

## Results and Discussion

We first determined the amino acid composition biases of each protein in the *S. cerevisiae* genome by dividing the observed frequency of a given amino acid by its proteomic

frequency. Next we grouped proteins and their amino acid composition biases using the Gene Ontology (GO) database, which can be used to cluster proteins by shared functions (see supplementary online material and Fig. S1). Genes sharing process GO annotations tend to be coexpressed (Chen and Xu 2004; Xiong et al. 2006) and are known a priori to be functionally related. Next we determined the amino acid biases of all groups defined by shared GO annotations compared with the proteomic average; the resulting amino acid composition biases of each 1,240 GO process annotation-defined functional group is a 20-dimensional vector we refer to as a “bias profile.” In determining this bias profile, we pooled the amino acids from all proteins before counting them rather than determining the average bias on a protein-by-protein basis, an approach that could potentially introduce artifacts if, for example, amino acid composition varied with protein size. In addition, by pooling protein sequence, each residue receives equal weight in the analysis, which is appropriate because substitutions occur on a residue-by-residue basis. However, preliminary analyses using GO biases calculated as the average of individual protein biases showed no qualitative difference.

We then clustered (Eisen et al. 1998) the bias profiles of the 221 GO groups containing at least 8 member proteins (Fig. 1), thereby excluding GO groups with  $\leq 7$  members, from which it is hard to extract statistically significant trends in amino acid bias. Choice of this cutoff decreased noise in the analysis, but it did not otherwise change the qualitative aspects of our results. Inspecting the hierarchical organization of GO groups clustered by similarity in bias profiles showed that functionally related GO groups have similar bias profiles. The relation between amino acid bias profile and protein function is preserved across great evolutionary distances (Fig. S2). Moreover, amino acid bias profiles are sufficient to predict the high-level cellular function of a protein pathway (Fig. S2). It should be recognized that individual proteins will occur in several GO groups; therefore, the GO groups cannot be considered strictly independent. To control for this, we randomly resampled the GO groups so that each protein could appear in at most one GO group. This did not qualitatively change the amino acid bias profiles of the 221 largest GO groups shown in Fig. 1.

To assess the relative contribution of potential selective pressures in addition to protein function to the variation in bias profiles, we more closely examined the clustering of amino acids in Fig. 1 by arranging the amino acid clustering tree in two dimensions. Inspection of this tree (Fig. 2a) showed that (with the exception of serine and histidine) it is comprised of two nonoverlapping clusters of amino acids, each of which has a common metabolic origin: 9 amino acids are generated by glycolysis, and 11

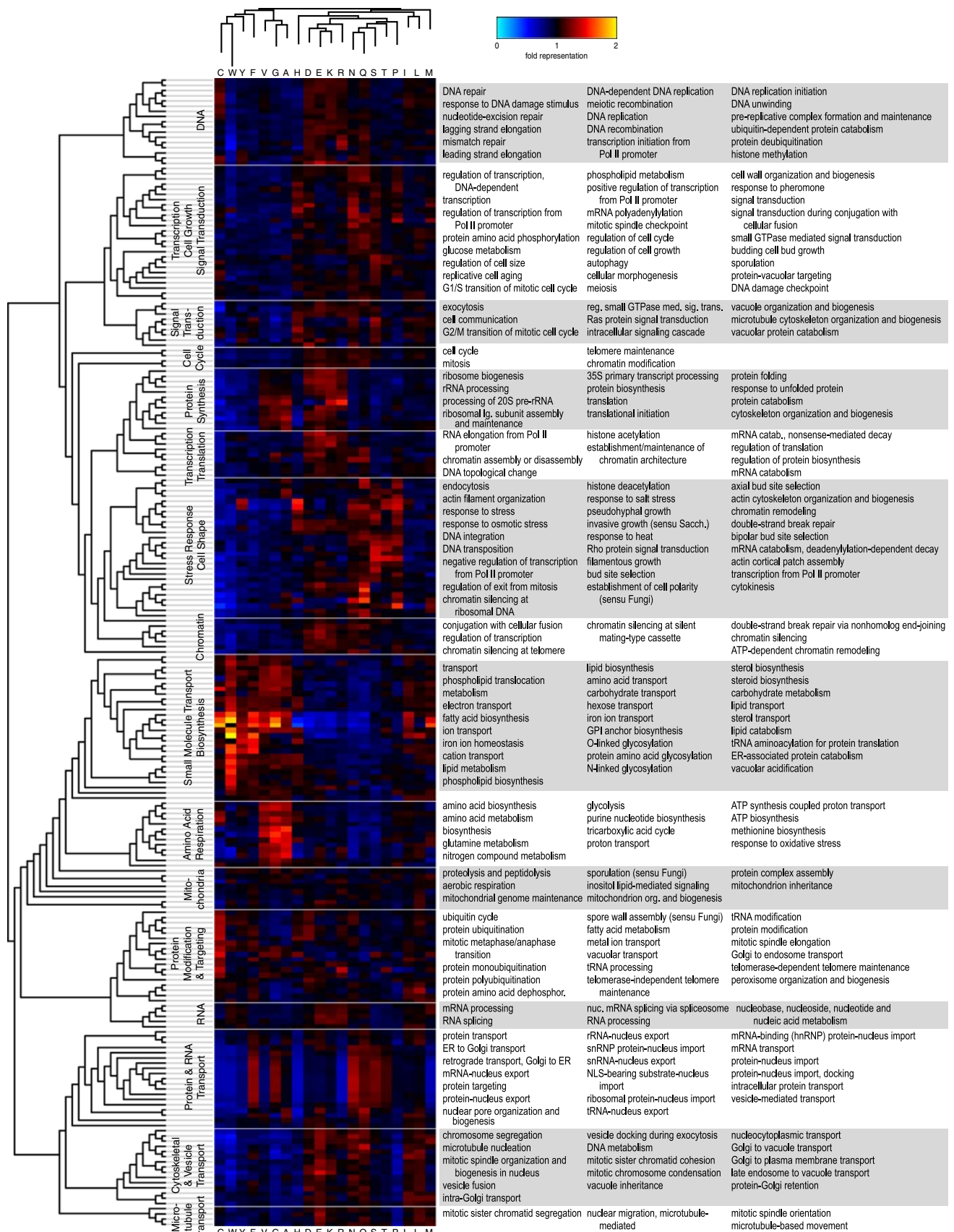
**Fig. 1** Amino acid composition bias profiles of the 221 largest GO groups. The fold abundance of each amino acid compared with the genomic average within the largest 221 GO groups is shown. *Red/yellow* indicates overrepresentation, and *blue/cyan* indicates underrepresentation. Rows are clustered by the similarity in bias profiles of each group, and columns are clustered by similarity in amino acid abundance across GO groups. Distance trees indicate the similarities of values in rows and columns, but they do not imply historic relations. Meta-GO groups that are inferred from groups within the GO clustering are labeled on the left

amino acids are generated by the tricarboxylic acid (TCA) cycle (Fig. 2b). This separation is highly statistically significant ( $p = 0.00046$  by Fisher's exact test, although it must be noted that synthetic rates of the individual amino acids are not strictly independent. Amino acid composition bias cannot arise passively as a function of the genetic code because there is no significant difference in the metabolic origin of amino acid produced by GC- versus AT-rich codons ( $p = 0.44$  by Fisher's exact test).

The coclustering of serine with amino acids derived from TCA suggests the existence of a cryptic biosynthetic link between serine and TCA. Indeed, such a link is supported by an alternate pathway for serine biosynthesis and is exemplified by experiments performed using *Escherichia coli* (Ravnikar and Somerville 1987) (and applicable to *Saccharomyces cerevisiae* given the presence of yeast orthologues of the relevant bacterial genes). The presence of histidine among the oxidatively derived amino acids may reflect its unique biosynthetic origin as the only amino acid derived from the pentose phosphate shunt. Moreover, histidine's association with oxidative amino acids is more sensitive to the organization of our data than that of serine (see Fig. S1).

The major separation of amino acids shown in Fig. 2a is congruent with amino acid metabolic source; however, this separation is also congruent with two additional physicochemical properties of amino acids: metabolic cost and polarity. First, the average metabolic cost of amino acids on the predominantly oxidative branch of the tree is significantly less than the average cost of amino acids on the reductive branch ( $p = 0.018$  by Mann-Whitney U-test). Metabolic costs constrain proteomic sequences in all domains of life; thus, it is not surprising that it also exerts an influence on amino acid composition bias across GO groups (Craig and Weber 1998; Swire 2007). Second, if one includes in the reductive branch the amino acids methionine, leucine, and isoleucine, then the tree significantly separates polar versus nonpolar amino acids ( $p = 0.0011$  by Fisher's exact test; Fig. 2a). The polarity of amino acids has also been previously cited as a major factor in amino acid use bias (Pascal et al. 2005; Nilsson et al. 2005). In identifying these factors, we tested 15 other amino acid properties, which exhibited no statistically



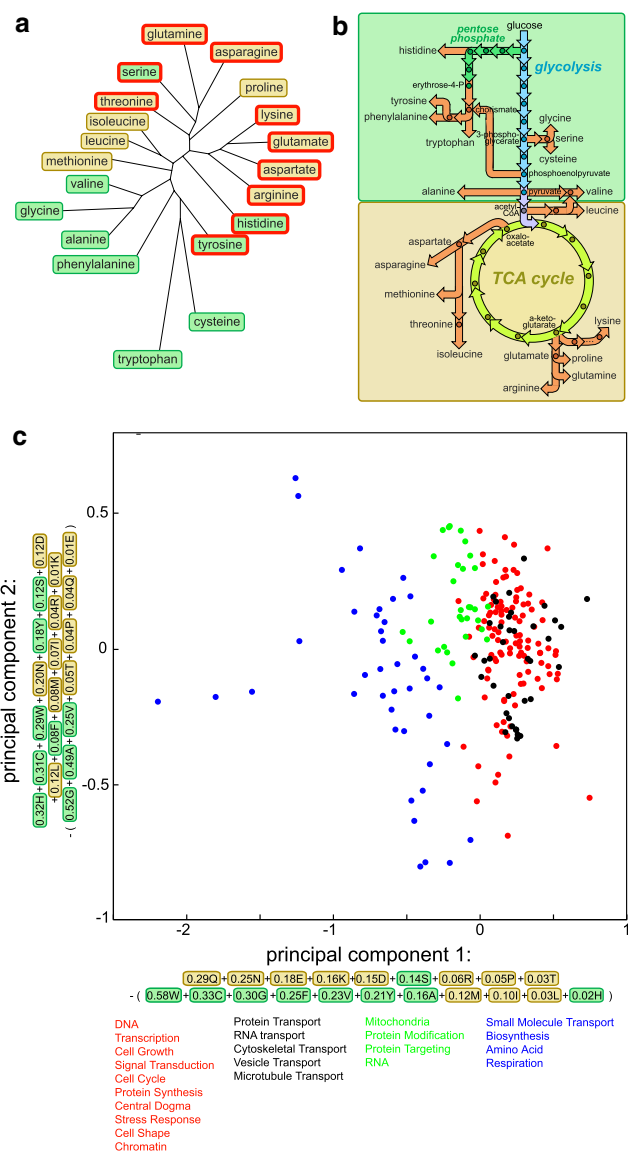


significant covariation across bias profiles, including van der Waals volume, pI, pK1, PK2, distribution of the amino acids across intracellular pools, and amino acid degree in the metabolic network.

We subjected the bias profiles of the 221 GO groups shown in Fig. 1 to principal component analysis (PCA), which yields uncorrelated eigenvectors (principal components) that explain the most variance in highly correlated, high-dimensional data sets, such as ours (Fig. 2c). The first principal component (PC1) explains the greatest variance (39%) in the data. Complementing what we observe in the compositional bias tree, PC1 encodes the metabolic source of amino acids. Overrepresentation of the oxidative amino acids contributes positively to PC1; conversely, overrepresentation of reductive amino acids decreases PC1. The average value of the normalized PC1 coefficients of oxidative amino acids is 0.040 and that of reductive amino acids is  $-0.22$  ( $p = 0.00061$  by Mann-Whitney U-test). PC1 also encodes amino acid metabolic cost and amino acid polarity because the metabolic cost of amino acids is highly correlated to their PC1 coefficients ( $p = 0.00094$  by Spearman rank correlation), and polar amino acids have significantly greater average PC1 coefficients than do nonpolar amino acids ( $p = 0.00043$  by Mann-Whitney U-test).

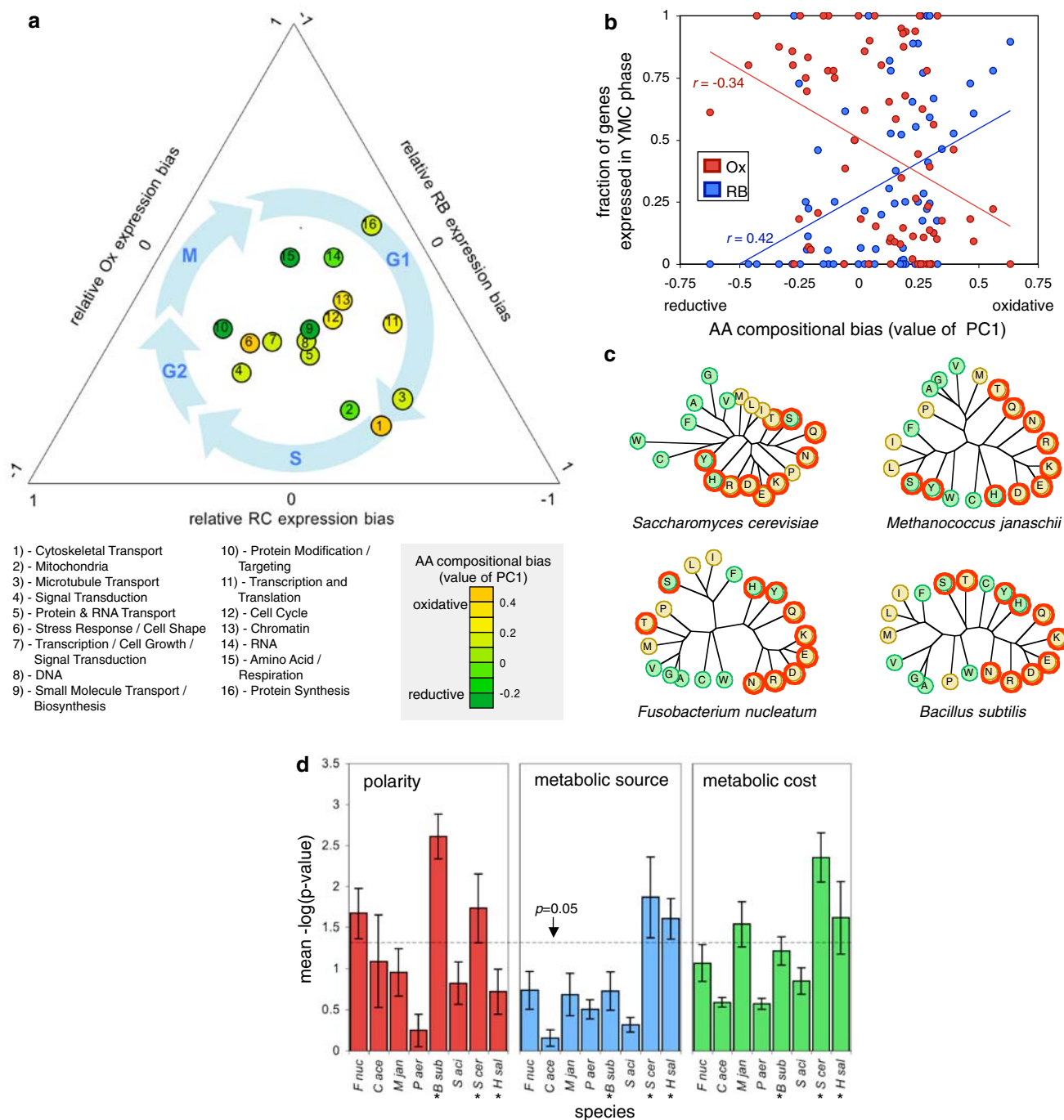
We next sought to determine which of these three factors was most important in amino acid bias variability. Our analyses determined that up to three amino acid properties, i.e., metabolic source, metabolic cost, and polarity, could be the predominant factors underlying amino acid composition bias across functionally related groups of proteins in yeast. It has been shown that the intracellular concentrations of amino acid precursors (e.g., acetate) fluctuate during the YMC (Tu et al. 2005; Ghosh and Chance 1964). Moreover, with the exception of asparagine, amino acids observed to have periodic intracellular concentrations are most abundant during the middle or at end of the metabolic phase that gives rise to their precursors (Hans et al. 2003; Wittmann et al. 2005). Therefore, we tested the hypothesis that composition bias may reflect an adaptive response to putative oscillations in intracellular amino acid concentrations by correlating composition bias (as encoded by PC1) to the YMC expression bias of GO groups, which was calculated as the fraction of genes expressed in each of the three metabolic phases: oxidative (Ox), reductive buildup (RB), reductive charging (RC) (Tu et al. 2005).

The fraction of periodic genes expressed in RB is positively correlated ( $r = 0.23$ ,  $p = 3.5 \times 10^{-14}$  by Spearman rank correlation) with PC1 across all 1,240 GO groups, whereas the fraction of genes expressed in Ox is negatively correlated to PC1 ( $r = -0.079$ ,  $p = 0.011$ ) (Fig. 3a, b). Therefore, GO groups predominantly expressed during Ox tend to be enriched in amino acids derived from the



**Fig. 2** Metabolic origin determines the greatest diversity in composition bias across GO groups. **a** The clustering of amino acids by similarity in bias profiles is shown in tree format. Amino acids derived metabolically from the oxidative portion of cellular respiration are yellow, whereas those derived from the reductive portion are green. Red outlines indicate polar amino acids. **b** Metabolic network of amino acid synthesis in yeast. **c** PCA of the bias profile across the 221 largest GO groups. Shown are the PC2 versus PC1 values of each GO group, along with the amino acid terms contributing to each PC. GO group points are colored according to membership in the listed meta-GOs

preceding reductive metabolic phase. Conversely, GO groups strongly expressed during RB and RC tend to be enriched in amino acids derived from Ox. These trends are stronger and even more statistically significant if one considers only those GO groups that are expressed the least uniformly across the YMC phases (Fig. 3b). There is no significant relation between RC bias and PC1 ( $p = 0.11$ ),



**Fig. 3** Amino acid abundances in periodic proteins follow the phases of their synthesis. **a** The same high-level functional groups seen in Fig. 1 in a ternary plot indicating their YMC phase expression biases. Color indicates PC1 value. Cell-cycle is shown for reference (blue). **b** The fraction of periodic genes expressed in the Ox (red) and RB (blue) phases are correlated with PC1 across the 75 GO groups exhibiting the most nonuniform YMC expression;  $p = 0.0042$  and  $0.00024$ , respectively, by Bonferroni-corrected Spearman rank correlation. **c** Composition bias trees (as in Fig. 2a). Yellow indicates oxidative amino acids, green indicates reductive amino acids, and red indicates polar amino acids. See Fig. S4 for remaining species. **d**

Significance of separation of amino acids by polarity, metabolic source, and metabolic cost across divisions of the composition bias trees. Bars indicate SEM across all trees divisions, leaving  $\geq 5$  amino acids on each “half.” Polarity, source, and cost were tested with Fisher’s exact, Fisher’s exact, and Mann–Whitney  $U$ -tests, respectively. Asterisks indicate facultative organisms. Obligate aerobes: *Pseudomonas aeruginosa* and *Sulfolobus acidocaldarius*. Obligate anaerobes: *Clostridium acetobutylicum*, *Fusobacterium nucleatum*, and *M. jannaschii*. Facultative species: *Halobacterium salinarium* and *Bacillus subtilis*

perhaps because the RC phase is intermediate to RB and Ox with respect to amino acid concentrations. The lag between the incorporation of TCA-derived amino acids in the reductive phase may also reflect a metabolic trade-off between diverting TCA intermediates from their role in energy production to their role as biosynthetic precursors.

Critically, of the three amino acid properties encoded by PC1, only metabolic source robustly correlates to expression timing. In particular, the correlations between the fractions of periodic genes expressed in RB and Ox, and the average metabolic cost of amino acids within GO groups, are, respectively,  $r = -0.0011$  ( $p = 0.97$ ) and  $r = 0.0064$  ( $p = 0.84$  by Spearman rank correlation). The fractions of periodic genes expressed in RB and Ox correlate better with the average hydrophobicity of amino acids within GO groups:  $r = -0.031$  ( $p = 0.32$ ) and  $r = 0.062$  ( $p = 0.044$ ), respectively, by Spearman rank correlation. By comparison, the fraction of oxidatively derived amino acids within GO groups is highly correlated to RB and Ox expression biases ( $r = 0.13$  ( $p = 0.000023$ ) and  $r = -0.11$  ( $p = 0.00041$ ), respectively, and even correlates with RC bias with marginal statistical significance:  $r = -0.062$  ( $p = 0.045$  by Spearman rank correlation). Statistics are given for all of these tests in Table S1. Therefore, the relation between PC1 and YMC expression bias is almost entirely caused by the metabolic source of amino acids.

These observations imply the following model: During evolutionary time, the amino acid content of periodically expressed proteins is fine-tuned by converting residues from amino acids that are less abundant to those that are more abundant at the time of translation. This optimization is advantageous provided that the global rate of protein translation decreases as a function of decreasing aminoacylated tRNA concentration or that translation error rate increases as a function of decreasing aminoacylated tRNA concentration (Akashi 2003). Accordingly, older groups of proteins would have sequences better optimized to their YMC phase than younger groups. Consistent with this prediction, we observed that older proteins (as measured by the number of their fungal orthologs) (Hsiang and Baillie 2005; Plotkin and Fraser 2007; Snel et al. 1999) show more negative correlation between Ox expression bias and sequence abundance of reductive amino acids (Fig. S3). This relation is statistically significant with  $r = -0.502$  ( $p = 0.0077$  by Spearman rank correlation).

Our model also implies that in the absence of predictable intrinsic fluctuations in amino acid concentrations, there would be no selective pressure to optimize protein sequences with respect to amino acid metabolic source. We repeated composition bias analysis using GO annotations and proteomic sequence data from eight ecologically and phylogenetically diverse microorganisms (see supplementary

online material) exhibiting either obligate aerobic, obligate anaerobic, or facultative metabolism. For this analysis, we included only organisms with experimental (rather than bioinformatic) evidence for their metabolic status and excluded a priori organisms we knew to grow symbiotically with other species or to be classified as facultatively an/aerobic on the basis of their randomly changing growth environment (such as *E. coli*, which switches between growth in and out of the gut). Fluctuations in amino acid concentrations in such organisms cannot correlate to their cell cycle-dependent expression timing.

For each of eight species satisfying these criteria, we generated composition bias trees and determined if these trees separated amino acids according to metabolic source, metabolic cost, or polarity (Figs. 3c, d; S4). The facultative organisms yeast and *Halobacterium salinarium* (Ruepp and Soppa 1996) showed strong variation in amino acid metabolic cost across GO groups; however, the obligate organisms *Methanocaldococcus jannaschii* and *Fusobacterium nucleatum* did as well. Similarly, some composition bias trees separated amino acids by polarity; however, this was uncorrelated to metabolic lifestyle for the tested organisms. However, none of the obligate organisms showed any statistically significant variation in amino acid bias with respect to metabolic source ( $p$  between 0.37 and 1 by Fisher's exact test for the deepest binary divisions of the composition bias trees), whereas the facultative organisms yeast and *H. salinarium* did ( $p = 0.00046$  and 0.028, respectively; the lower significance in the latter species may reflect a dilution of the amino acid optimization signal because of nucleotide-level GC-content pressure) (Singer and Hickey 2000).

Although the phylogenetic and ecologic complexity of aerobic/reductive metabolism across microorganisms means that this restricted set of observations cannot lead to a definite conclusion about all microorganisms, for this group of organisms facultative metabolism is required for a predictive relation between amino acid composition bias and amino acid metabolic source. In general, we would predict that under any conditions in which fluctuations in amino acid prevalence occur predictably with respect to protein expression timing, there is directional pressure on amino acid substitutions that are not specifically constrained by protein function. This effect is predicted to be timescale independent: Previous work has shown that changes in global atmospheric atomic composition may have exerted selective pressure on proteome sequence during geologic timescales (Acquisti et al. 2007). Our examination of amino acid composition bias implies that the influence of ultradian bioenergetics, a previously unappreciated selective pressure, has exerted a subtle but widespread effect on proteome sequence during evolutionary timescales. Although not explicitly designed to test



the selectionist hypothesis of molecular evolution, our results, along with the prevalence of evidence for the optimization of genomic sequences under numerous metabolic constraints, are inconsistent with the view that substitution events are predominantly selectively neutral.

**Acknowledgements** Thanks to Boris Magasanik, David Liu, and Alan Saghatelian for constructive comments on the manuscript. Funding was provided by a Merck-Wiley Fellowship (B. L. D.) and the National Institute of General Medicine Sciences (S. L. S.).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Acquisti C, Kleffe J, Collins S (2007) Oxygen content of transmembrane proteins over macroevolutionary time scales. *Nature* 445:47–52
- Akashi H (2003) Translational selection and yeast proteome evolution. *Genetics* 164:1291–1303
- Akashi H, Gojobori T (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci USA* 99:3695–3700
- Alves R, Savageau MA (2005) Evidence of selection for low cognate amino acid bias in amino acid biosynthetic enzymes. *Mol Microbiol* 56:1017–1034
- Baudouin-Cornu P, Surdin-Kerjan Y, Marliere P, Thomas D (2001) Molecular evolution of protein atomic composition. *Science* 293:297
- Chen Y, Xu D (2004) Global protein function annotation through mining genome-scale data in yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res* 32:6414–6424
- Craig CL, Weber RS (1998) Selection costs of amino acid substitutions in ColE1 and ColIa gene clusters harbored by *Escherichia coli*. *Mol Biol Evol* 15:774–776
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868
- Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. *Science* 155:279–284
- Ghosh A, Chance B (1964) Oscillations of glycolytic intermediates in yeast cells. *Biochem Biophys Res Commun* 16:174–181
- Hans MA, Heinzle E, Wittmann C (2003) Free intracellular amino acid pools during autonomous oscillations in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 82:143–151
- Hsiang T, Baillie DL (2005) Comparison of the yeast proteome to other fungal genomes to find core fungal genes. *J Mol Evol* 60:475–483
- Lobry JR, Gautier C (1994) Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic Acids Res* 22:3174–3180
- Mazel D, Marliere P (1989) Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. *Nature* 341:245–248
- Nilsson J, Persson B, von Heijne G (2005) Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes. *Proteins* 60:606–616
- Pascal G, Medigue C, Danchin A (2005) Universal biases in protein composition of model prokaryotes. *Proteins* 60:27–35
- Perlstein EO, de Bivort BL, Kunes S, Schreiber SL (2007) Evolutionarily conserved optimization of amino acid biosynthesis. *J Mol Evol* 65:186–196
- Plotkin JB, Fraser HB (2007) Assessing the determinants of evolutionary rates in the presence of noise. *Mol Biol Evol* 24:1113–1121
- Porter TD (1995) Correlation between codon usage, regional genomic nucleotide composition, and amino acid composition in the cytochrome P-450 gene superfamily. *Biochim Biophys Acta* 1261:394–400
- Ravnikar PD, Somerville RL (1987) Genetic characterization of a highly efficient alternate pathway of serine biosynthesis in *Escherichia coli*. *J Bacteriol* 169:2611–2617
- Richard P (2003) The rhythm of yeast. *FEMS Microbiol Rev* 27:547–557
- Ruepp A, Soppa J (1996) Fermentative arginine degradation in *Halobacterium salinarum* (formerly *Halobacterium halobium*): genes, gene products, and transcripts of the arcRACB gene cluster. *J Bacteriol* 178:4942–4947
- Seligmann H (2003) Cost-minimization of amino acid usage. *J Mol Evol* 56:151–161
- Singer GA, Hickey DA (2000) Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Mol Biol Evol* 17:1581–1588
- Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco, CA
- Snel B, Bork P, Huynen MA (1999) Genome phylogeny based on gene content. *Nat Genet* 21:108–110
- Swire J (2007) Selection on synthesis cost affects interprotein amino acid usage in all three domains of life. *J Mol Evol* 64:558–571
- Tekaia F, Yeramian E (2006) Evolution of proteomes: Fundamental signatures and global trends in amino acid compositions. *BMC Genomics* 7:307
- Thiolouse J, Lobry JR (1995) Co-inertia analysis of amino-acid physico-chemical properties and protein composition with the ADE package. *Comput Appl Biosci* 11:321–329
- Tu BP, Kudlicki A, Rowicka M, McKnight SL (2005) Logic of the yeast metabolic cycle: Temporal compartmentalization of cellular processes. *Science* 310:1152
- Wittmann C, Hans M, van Winden WA, Ras C, Heijnen JJ (2005) Dynamics of intracellular metabolites of glycolysis and TCA cycle during cell-cycle-related oscillation in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 89:839–847
- Xiong J, Rayner S, Luo K, Li Y, Chen S (2006) Genome wide prediction of protein function via a generic knowledge discovery approach based on evidence integration. *BMC Bioinformatics* 7:268